METHOD AND ITS APPARATUS FOR CLASSIFYING DEFECTS

Background of the Invention

The present invention relates to a method for automatic classification of data. Particularly, the present invention relates to a method for automatic classification of defects occurred on the surface of a semiconductor electronic circuit board, a printed circuit board, a liquid crystal display board or the like according to a detected image, an EDX detection spectrum, or the like.

Recently, methods for automatic classification by detecting an image of a defect portion have been developed in order to quickly grasp the situation of the defect occurred on the surface of a semiconductor electronic circuit board or the like and to monitor the number of occurrences per each type of defect.

For the automatic classification of images, various methods are conventionally studied in a field of pattern recognition.

One of conventional methodologies is a method called the learning type classification. According to this methodology, a teacher image is collected in advance and learned to optimize a classification apparatus (neural network, etc.). A learning type classification apparatus has a possibility that it can

classify with flexibility in accordance with the request made by a user but has a disadvantage that it cannot be used substantially at the startup of a production process because it is generally necessary to

5   collect a large volume of teach data so as to obtain good performance. It is known that, when a small volume of teach data alone is used, a phenomenon of excessive conformity of learning with the teach data, which is called overlearning, occurs, resulting in

10   degradation in performance.

There is also another conventional methodology called a rule-based classification method. According to this methodology, a characteristic amount is extracted from an image to be classified, and the

15   value of the characteristic amount is judged according to the "if-then" rule incorporated into the system to classify a defect into one of classes. A rule-based classification apparatus cannot respond flexibly to the request by the user because a class rule for

20   classification is fixed but has an advantage that it can be used from the startup of the production process because teach data is not required.

The above-described rule-based classification apparatus and the learning type classification appara-

25   tus may be used together as one method. An example of such a method is disclosed in Japanese Patent Laid-Open Publication No. 2001-135692. Specifically, a defect is classified into a previously incorporated fixed number

of classes (called the "core classification") by the rule-based classification apparatus called the "core classifier" and further classified into an arbitrary number of "low-order classification" by the learning type classification apparatus called the "particular applicable classifier" which is associated with the core classification.

The example disclosed in the above-described patent publication uses the core classifier and can conduct the core classification from the startup of the process without necessity of collecting an amount of teach data. If classification in further detail is required, the classification can be made by the learning type "particular applicable classifier".

The above-described prior art needs to decide previously a classification model combining the rule-based classification apparatus and the learning type classification apparatus. But, it is generally very hard to determine an optimum classification model in advance, possibly resulting in degradation in performance because the classification model is inadequate. Description will be made by examples below.

Fig. 2 to Fig. 4 show three types of classification models for classification of defects into four classes of an on-the-film foreign material, a below-the-film foreign material, a recess and a pattern defect. It will be described below that an optimum

classification model is variable depending on a distributed state of defects.

Fig. 2 shows an example of a single layer classification model. A rule-based classification apparatus 21 corresponds to Section 1 and classifies into four classes of an on-the-film foreign material 22, a below-the-film foreign material 23, a recess 24 and a pattern 25.

The rule-based classification apparatus is superior to the learning type classification apparatus on the point that it can deliver stable performance as long as a designed rule adequately matches the target to be classified.

In a field of classification of defects, the causes of defects have become diverse with a technological evolution of the production process, and the classes for classification of defects have also varied accordingly. Therefore, it is hard to classify the defects of products, which are produced through different production processes, by using universal defect classification classes, and it must be said that a possibility of applicability of the rule previously assumed for a prescribed production process by a designer to the products produced by a different production process is very low. In this connection, the classification classes such as handwritten numeric recognition classification, etc. are considerably different from the setting determined at the time of

designing.

Fig. 3 shows a double-layered classification model. A first layer's rule-based classification apparatus 31 classifies into three classes of a foreign material 32, a recess 33 and a pattern 34, and a second layer's learning type classification apparatus 35 further classifies the foreign material 32 into two classes of an on-the-film foreign material 36 and a below-the-film foreign material 37.

According to the example of the single layer classification model shown in Fig. 2, a detected defect is classified into one of the four classes desired by the user, while according to the model shown in Fig. 3, a probability of applicability of the designed rule can be made high on the point that the classification into three classes or any of them is conducted. Meanwhile, when the classification into the on-the-film foreign material or the below-the-film foreign material is conducted by the second layer's learning type classification apparatus, a possibility capable of classifying with reliability higher than the rule-based classification assumed by the designer becomes high in a condition that teach data on the on-the-film foreign material and the below-the-film foreign material is adequately large.

Fig. 4 shows a double-layered classification model of a type different from that shown in Fig. 3. The first layer classifies into three classes of a

foreign material, a recess and a pattern, and the
second layer further classifies the foreign material
into two classes of an on-the-film foreign material and
a below-the-film foreign material. A classification
5   apparatus corresponds to Section 1 and Section 2 of the
classification model. Here, it is assumed that Section
1 is a rule-based classification apparatus and Section
2 is a learning type classification apparatus.

In the example shown in Fig. 4, Section 2 is
10   different from that of the classification apparatus
shown in Fig. 3 and seems something different. The
learning type classification apparatus of Section 2 is
a learning type classification apparatus, which
classifies the defect classified as a foreign material
15   by the classification apparatus of the first layer into
an on-the-film foreign material, a below-the-film
foreign material or a pattern defect. The recess and
the pattern defect can be separated with high
reliability by the rule-based classification apparatus,
20   but there is a possibility that the foreign material
only is separated with high classification performance
as compared with the model shown in Fig. 3 in a
situation that it cannot be separated from the pattern
defect.

25      Besides, a big difference of the
classification model shown in Fig. 4 from the
classification trees shown in Figs. 2 and 3 is that it
is a classification model different from a hierarchical

relationship (semantic classification model) in terms of a classification concept the user has. The foreign material and the pattern defect are of exclusive classes from each other in terms of the user's

5   classification concept, and the pattern defect cannot be located below the foreign material. But, the classification model achieving the maximum classification performance and the user's conceptual classification model can be independent mutually except

10   when the bottom layer is a class finally classified by the user. It also suggests that it is hard for the user to determine an optimum classification model.

It can be said from the above that the optimum classification model for an automatic defect

15   classification problem is variable depending on the problem setting. And, this problem setting (a level of adequacy of the rule by a designer, a possibility of collecting learn data, etc.) cannot be assumed in advance, so that an optimum classification model cannot

20   be determined in advance either. As a result, there occurs a problem that the classification performance drops because the classification model is not optimum.

To achieve the maximum performance, the classification tree automatically provides an inherent

25   and optimum classification model in response to a user's defect classification request, which is variable depending on the users, thereby improving the classification performance. Besides, it eliminates the

necessity of manual setting of the classification model.  It is hard for the user to previously determine the optimum classification model because it does not always match the conceptual classification model

5    (generally called the semantic gap).


Summary of the Invention

The present invention remedies the above described disadvantages of the prior art and relates to a method and its apparatus for classifying defects,

10   which are capable of improving classification performance by automatically providing an inherent and optimum classification model in response to a defect classification request which is variable depending on users or making it possible for the user to easily

15   select it.

Specifically, the invention first generates one or plural potential classification models.  For example, the potential classification model is comprised of a hierarchical structure of the rule-based

20   classification apparatus and the learning type classification apparatus as shown in Fig. 2, Fig. 3 or Fig. 4.  It is assumed that the potential classification model is set by the user through the screen, set by using teach data or set automatically as described

25   in detail later.  When plural classification models are generated, an index (hereinafter referred to as the model likelihood) about the adequacy of each model is

calculated. In other words, the model likelihood is an index indicating a degree of reliability that the likelihood calculated by each potential classification model can be trusted.

5       Then, the class likelihood of each potential classification model is calculated. According to the model likelihood, the class likelihoods calculated by the plural classification models are combined to make classification. For the combination of the class
10 likelihoods, the likelihoods are combined with importance relatively attached to a high model likelihood or a class likelihood calculated by a reliable model and without importance relatively attached to a low model likelihood or a class
15 likelihood calculated by an unreliable model.

      Thus, an inherent and optimum classification model can be set easily or can be provided automatically in response to a defect classification request variable depending on users, and it becomes
20 possible to improve classification performance.

      These and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the
25 accompanying drawings.

Brief Description of the Drawings

      Fig. 1A and Fig. 1B are diagrams each showing

a flow of generation of a classification model
according to a first embodiment;

Fig. 2 is a diagram showing an example of the
classification model;

Fig. 3 is a diagram showing another example
of the classification model;

Fig. 4 is a diagram showing another example
of the classification model;

Fig. 5 is a diagram illustrating a
classification model according to the first embodiment
of the invention;

Fig. 6 is a diagram illustrating another
classification model according to the first embodiment
of the invention;

Fig. 7 is a diagram illustrating another
classification model according to the first embodiment
of the invention;

Fig. 8 is a block diagram showing a schematic
structure of a defect review apparatus according to the
first embodiment of the invention;

Fig. 9A to Fig. 9C are front diagrams of
display screens showing user interfaces according to
the first embodiment, where Fig. 9A shows a rule-based
classification screen, Fig. 9B shows a learning type
classification screen, and Fig. 9C shows a recipe
setting screen;

Fig. 10A and Fig. 10B are graphs each showing
a distribution of a characteristic amount of defects;

Fig. 11 is a diagram showing a classification model;

Fig. 12 is a diagram showing a flow of processing for classification;

5 Fig. 13 is a diagram showing likelihood function of a rule type classification apparatus;

Fig. 14 is a diagram illustrating a Mahalanobis distance;

Fig. 15 is a diagram showing an example of a 10 learning type classification model;

Fig. 16 is a diagram showing another example of the learning type classification model;

Fig. 17 is a diagram showing an example of a rule type classification model; and

15 Fig. 18 is a diagram showing another example of the rule type classification model.

Description of the Embodiments

The first embodiment of the invention will be described with reference to the accompanying drawings.

20 First, the present invention detects a defect of a sample by a defect inspection apparatus using an optical microscope or an electron microscope. Here, the sample to be inspected includes a semiconductor electronic circuit board, a printed circuit board, an 25 FPD, a liquid crystal display board and the like. For classification of defects, a defect portion caused on the surface of such a board is detected by the defect

inspection apparatus using the optical microscope or the electron microscope, position coordinate information on the defect detected on the board is stored, the sample is positioned based on the stored

5   position coordinate information on the defect so that the defect on the sample falls in the detection field of view of a review apparatus, an image of the defect is obtained by the review apparatus, and the obtained image of the defect is processed.

10          Fig. 8 shows a schematic structure of the defect review apparatus according to the invention.  In Fig. 8, 80 denotes an electron microscope which deflects an electron beam 83 by a deflector 84 according to the position coordinate information on the

15  defect detected by the defect inspection apparatus to scan and irradiate a semiconductor substrate 82, which is subject to the inspection and positioned within a housing 81, with the electron beam 83, and detects by a detector 85 a secondary electron produced from the

20  semiconductor substrate 82 in synchronization with the deflected signal from the deflector 84.  The detected secondary electron is converted into an electric signal, which is then converted into a digital image signal by an A/D converter 86, and it is input as a

25  defect image signal to a processing unit 87.  The processing unit 87 analyzes the input digital defect image signal by the software processing to classify automatically.  The processed result and the detected

image are stored in a storage device 88.

Then, a method of classifying the defect by processing the defect image obtained by the review apparatus will be described.

5    (1)   Outline of defect classification

First, the defect classification according to the invention is described briefly with reference to Fig. 1A and Fig. 1B.  The defect classification comprises two steps of (i) previous generation of a

10   classification model before the execution of classification processing and (ii) execution of classification.

(i)   Generation of classification model

Numeral 11 denotes a state before the

15   generation of a classification model.  M1, M2 and M3 in the drawing denote classes (hereinafter referred to as the main classes) which are previously provided by the system of the present invention, and U1 through U4 denote classes (hereinafter referred to as the user

20   classes) determined by the user.  The user determines the user classes manually or by using teach data on the setting screen.  Here, it is not determined previously which user class is located below which main class. Meanwhile, for example, at least one model having a

25   combination of main classes and user classes each indicated by 12 and 13 is generated.  The model

generation may be set manually on the setting screen or made automatically by the method to be described layer. In the automatic generation, the correlation between the main class likelihood-calculated result by the

5 rule-based classification apparatus and the user class taught by the user is analyzed.

Here, it will be described with reference to Fig. 10A and Fig. 10B that an appropriate model is variable depending on defect data and, therefore, there

10 is a possibility of improving the classification performance by assuming not a single classification model but plural classification models.

A graph 101 in Fig. 10A and a graph 102 in Fig. 10B each show an example of the distribution of a

15 characteristic amount. In the graph 101 in Fig. 10A, the "foreign material", the "recess" and the "pattern" are distributed to adequately separate from one another, so that it is expected that the classification model combining the rule-based classification apparatus

20 31 and the learning type classification apparatus 35 as shown in Fig. 3 is suitable. When it is assumed that likelihood functions $L_{foreign\ material}(f1,\ f2)$, $L_{pattern}(f1,\ f2)$ and $L_{recess}(f1,\ f2)$ according to the rule-based classification apparatus are defined as follows, the

25 rule-base type likelihood becomes a very effective criterion for classification.

When $L_{foreign\ material}(f1,\ f2)$ : $f1 < Th0$, the likelihood of the foreign material is large;

when $L_{pattern}$(f1, f2) : f1>Th0 and f2>Th1, the likelihood of the pattern is large; and

when $L_{recess}$(f1, f2) : f1>Th0 and f2<Th1, the likelihood of the recess is large.

5        Meanwhile, in the distributions shown in the graph 102 of Fig. 10B, the rule-based classification apparatus 31 configured as shown in Fig. 3 comes to have a high possibility of classifying not only the "foreign material" 32 but also the "pattern" 34 into a

10  large number of "foreign material" 32. Therefore, when the "foreign material" 32 is further subdivided by the learning type classification apparatus 35, the classification model having the combination of the rule-based classification apparatus 31 and the learning

15  type classification apparatus 35 shown in Fig. 3 is not optimum necessarily. Conversely, when the performance of the learning type classification apparatus is adequately reliable, the classification model having the combination of a rule-based classification

20  apparatus 41 and a learning type classification apparatus 45 shown in Fig. 4 is considered more appropriate.

        Actually, the optimum classification model is different depending on data. For example, for the

25  defect data having the characteristic amount distributed as shown in the graph 102 of Fig. 10B, the classification model shown in Fig. 4 is more adequate than the classification model shown in Fig. 3 (the

classification model shown in Fig. 3 cannot appropriately classify the pattern defect of Fig. 10B), but for the defect data having the characteristic amount distributed as shown in the graph 101 of Fig.

5   10A, the classification model shown in Fig. 3 can be expected having higher classification reliability than the classification model shown in Fig. 4.

As described above, the suitable classification model is different depending on defect

10  data, so that there is a possibility of making it possible to improve the classification reliability by using plural classification models together than a particular classification model is adopted.  The present invention generates plural models, calculates a

15  criterion (model likelihood) which indicates the adequacy of a model about which model shall be used for defect data and classifies according to the criterion.

(ii)   Execution of classification

Defects are classified according to at least

20  one classification model generated as described above. When plural classification models are generated, they are used in combination to execute classification.  In other words, the likelihoods belonging to individual classes under the above-described individual models 12,

25  13 are calculated (14, 15).  Then, there is a problem of how to calculate a final class likelihood from the likelihoods calculated independently under the

individual models. According to the present invention, the model likelihoods of the individual models 12, 13 are calculated as a criterion indicating the adequacy of the pertinent model per se (16, 17), and a final

5 class likelihood (combination likelihood) is calculated considering the model likelihoods (18). The method of calculating the model likelihood and the method of calculating the final class likelihood will be described later.

10 (2) Details of defect classification

Specific examples of defect classification according to the invention will be described in detail below.

(i) Generation of classification model

15 First, a classification model is previously generated before the classification processing is executed. Fig. 5 is a diagram schematically showing the relation between main classes and user classes before the classification model is generated.

20 For the classification model of this embodiment, it is assumed that a classification model is comprised of two layers of a high-order section 1 which is comprised of a rule-based classification apparatus 51, and low-order sections 2 to 4 which are

25 comprised of learning type classification apparatuses 55 to 57. The rule-based classification apparatus 51

of the high-order section 1 calculates likelihoods
belonging to three main classes of a "foreign material"
52, a "recess" 53 and a "pattern" 54 according to the
if-then rule.  Meanwhile, when it is assumed that a

5  defect is classified into a certain main class, the
learning type classification apparatuses 55 to 57 of
the low-order sections 2 to 4 calculate likelihoods of
the classes (user classes 1 to 4:58-1 to 58-4) which
are defined as low-order classes of the main class.

10  The relationship between the main class and the user
classes is not defined in Fig. 5, but it is assumed in
this embodiment that the defect is finally classified
into any of the user classes 1 to 4.

Then, a method for generation of a classifi-

15  cation model will be described.  The classification
model shown in Fig. 6 or Fig. 7 is generated from the
state shown in Fig. 5 by the method described below.

User interfaces are shown in Fig. 9A through
Fig. 9C.

20  Fig. 9A shows a main class screen, where 91
denotes a rule-based classification screen.  A list 911
of rule-base type classes and a list 912 of defect
images showing that the likelihood of each rule-base
type class becomes maximum, namely the most typical of

25  the pertinent class, are shown.

Meanwhile, Fig. 9B shows a user class screen,
where 92 denotes a screen (hereinafter referred to as
the learning type classification screen) for teaching

of classes by the learning type classification
apparatus and is comprised of a list 921 of learning
type classes and a list 922 of defect images. The user
sets the class of a defect to be classified on the

5   learning type classification screen 92 as a learning
class and teaches the each defect image 922 to any of
the learning type classes 921 by a drag and drop
operation.

Then, a button "automatic link" 923 icon for

10  instruction of automatic generation of a classification
model is clicked to internally calculate the structure
of the classification model.

To generate the classification model, a
correlation between the main class and the user class

15  is analyzed from the taught result by the user and the
likelihood calculated result by the rule-based
classification apparatus, and a classification model
combining a rule-based classification apparatus 601 or
701 and learning type classification apparatuses 602,

20  603 or 702, 703 as shown in Fig. 6 or Fig. 7 is
generated. Specifically, it is assumed that (1) a
correlation between a main class $Mi$ and a user class $Uj$
is high when there are many defects which are taught as
the class $Uj$ by the learning type classification

25  apparatus among the defects which are determined by the
rule-based classification apparatus that the likelihood
of the main class $Mi$ is maximum, or (2) a correlation
between the main class $Mi$ and the user class $Uj$ is high

when the defects classified into the main class Mi are many among the defects taught as the user class Uj, and a hierarchical relationship is set between the main class and the user class which have a high correlation.

5 For example, the above (1) and (2) can be formularized as follows.


[Classification model generation rule 1]

When $N_{Uj}/N_{Mi}$ is $th_1$ or more or $N_{Mi}/N_{Uj}$ is $th_2$ or more, the user class Uj is determined to be a lower

10 class of the main class Mi,

where:

$N_{Mi}$: the number of defects that the likelihood of the main class Mi is determined to be maximum by the rule-based classification apparatus;

15 $N_{Uj}$: the number of defects that the likelihood of the user class Uj is determined to be maximum by the learning type classification apparatus; and

$N_{Mi \cap Uj}$: the number of defects that the likelihood of the main class Mi is determined to be maximum by the

20 rule-based classification apparatus and the likelihood of the user class Uj is determined to be maximum by the learning type classification apparatus.

A classification model suitable to the obtained teach data can be generated automatically by

25 the above-described method.

A method of generating plural classification models when the suitable classification model is not

necessarily single with respect to the distribution of the obtained teach data will be described.

In the example of Fig. 5, it is assumed that the user classes 1, 2 (58-1, 58-2) have a high correlation with the main class "foreign material" 52 and the user class 4 (58-4) has a high correlation with the main class "pattern" 54. It is also assumed that there are five samples of the user class 3 (58-3) among which four samples are with the maximum likelihood of the main class "pattern" 54 and one sample is with the maximum likelihood of the main class "foreign material" 52. According to the classification model generation rule 1, most (4/5=80%) of the defects belonging to the user class 3 (58-3) are with the maximum likelihood of the main class "pattern" 54, so that it seems adequate to set the user class 3 (58-3) to be below the main class "pattern" 54 but, actually it cannot be said with certainty that the likelihood of the main class 3 "pattern" 54 becomes maximum for most of the defects belonging to the user class 3 (58-3) because the amount of data is too small. Therefore, for example, when the amount of data is little, assumable plural models are generated by the next classification model generation rule 2 in addition to the classification model generation rule 1. According to the classification model generation rules 1, 2, the classification model combining the rule-based classification apparatus 601 or 701 and the learning type classification apparatuses

602, 603 or 702, 703 as shown in Fig. 6 or Fig. 7 is generated in the above example.

[Classification model generation rule 2]

When the number of defects $N_{Uj}$ belonging to

5  the user class Uj is a fixed level or below, plural classification models are generated as follows.  It is assumed that, when defects belong to the user class Uj and the number of defects that the likelihood belonging to the main class Mi becomes maximum is $N_{Mi \cap Uj}$, a main

10  class having the maximum value of $N_{Mi \cap Uj}$ and a main class having the second maximum value are determined to be Mmax1 and Mmax2 respectively.  A model having the user class Uj determined to be below the Mmax1 and a model having the user class Uj determined to be below the

15  Mmax1 and the Mmax2 are generated.

The method for automatic generation of a classification model for the obtained teach data was described above.  It is also possible to manually determine a model or manually modify the automatically

20  generated model.  On the user interface screens of Fig. 9A through Fig. 9C, a "recipe setting screen" button 924 icon shown on the learning type classification screen 92 is clicked to jump to a recipe setting screen 93 of Fig. 9C, on which the configuration of the

25  generated classification model is checked, and the user can also manually change the automatically generated model as required on the recipe setting screen 93.

Fig. 9C shows only one classification model, but it may be configured to calculate, display and modify plural possible classification models.

It was described that the two classification
5  models were comprised of the hierarchical structure of the rule type classification apparatus and the learning type classification apparatus. Using the same classification model configuration method, a classification model comprising a combination of the
10  plural learning type classification apparatuses (1501 to 1503 or 1601 to 1603) as shown in Fig. 15 or Fig. 16 can be configured as another structure. As still another configuration, a classification model may be configured by combining the plural rule type
15  classification apparatuses (1701 to 1703 or 1801 to 1803) as shown in Fig. 17 or Fig. 18.


(ii)  Execution of classification

Then, a process for execution of classification will be described in detail. Fig. 12
20  shows a flow of processing for the execution of classification. First, calculation of a likelihood of each classification model will be described with reference to Fig. 12.


STEP 1:  Calculation of likelihood in each
25  classification model (1201)

This step comprises three steps including

calculation of main class likelihood, calculation of
user class likelihood and calculation of rule learning
integration likelihood from them.

STEP 1.1:  Calculation of main class likelihood (1202)

5          The rule-based classification apparatus does
not classify into a particular class but calculates the
likelihood of each class on the basis of fuzzy logic
(fuzzy if-then rule).  The likelihood is an index
indicating a degree of probability of belonging to the

10  class and falls in a value range of 0 to 1.  It
indicates that the closer to 1, the higher the degree
of probability of belonging to its class.  A function
for calculation of the likelihood from a value of a
characteristic amount is called the likelihood function

15  and provided by the system.  An example of a main class
likelihood function is shown in Fig. 13.  In the
drawing, 131 denotes a likelihood function of a foreign
material, 132 denotes a likelihood function of a
pattern and 133 denotes a likelihood function of a

20  recess.  The likelihood to each class is calculated
from a value of a characteristic f1 by 131, 132 and
133.

STEP 1.2:  Calculation of user class likelihood (1203)

          A method for calculation of the user class

25  likelihood by the learning type classification
apparatus is shown in Fig. 14.  In the drawing, 1401

denotes an example of the likelihood function for
calculation of the user class likelihood.  As the
likelihood function, for example, the Mahalanobis
distance in a characteristic space may be used.  The
5   Mahalanobis distance is a distance determined by
assuming that distributions 141, 142 of teach data on
individual classes comply with (generally multi-
dimensional) normal distributions 143, 144 in a
characteristic amount space and standardizing a
10   distance from the center of each normal distribution to
the characteristic amount vector of a pertinent defect
by the dispersion of a normal distribution (covariance
matrix).  Here, the likelihood of the learning type
class can be calculated as follows.

15   (Likelihood belonging to learning type class i) =
$D_i/\Sigma D_i$
where, $D_i$ is the Mahalanobis distance between a
learning type class $D_i$ and a characteristic amount
vector of defects to be classified.
20           A difference from the likelihood function of
the main class likelihood is that the likelihood
function is determined by teach data.

STEP 1.3:  Calculation of rule learning integration
likelihood (1204)
25           Then, class likelihood combining the
calculated main class likelihood and the user class

likelihood is calculated (hereinafter referred to as the rule learning integration likelihood). The user class likelihood is likelihood defined on the assumption that classification into a high-order main

5  class is made. Therefore, even if the user class likelihood has a high value, it is necessary to define so that the rule learning integration likelihood of the user class has a low value when its high-order main class likelihood has a very low value.

10      For example, the rule learning integration likelihood can be calculated as an average of the load of the main class likelihood and the user class likelihood.

(Rule learning integration likelihood) =

15      $\alpha *$ (main class likelihood) $+ (1-\alpha) *$ (user class likelihood)

where, $\alpha$ denotes a degree of placing emphasis on the main class likelihood and may be determined descendingly on the basis of the empirically obtained

20  reliability of the rule type classification apparatus and the learning type classification apparatus. As another calculation method, it is conceivable that the main class likelihood and the user class likelihood are added, or the like.

25      The above-described step is executed for each classification model to determine a class likelihood under each classification model.

STEP 2: Calculation of model likelihood (1205)

Then, for each generated classification model, model likelihood is calculated as an index indicating the adequacy of the model.

Among the plural models generated, an optimum classification model is actually different depending on each defect. For example, it is assumed that two types of models shown in Fig. 3, Fig. 4 are generated. When it is apparent from the value of a characteristic amount that the defect to be classified is a pattern defect and it is not necessary to particularly calculate the likelihood by the learning type classification apparatus, it is conceivable that the model of Fig. 3 is more advantageous than the model of Fig. 4 which has a possibility of misclassification into a foreign material. Meanwhile, when it is not apparent from the value of the characteristic amount whether the defect is a pattern defect or a foreign material, it is considered more advantageous to classify by the learning type classification apparatus according to the model of Fig. 4. Thus, the optimum classification model is actually different for each defect.

For example, the model likelihood can be defined for each defect as follows:

Model likelihood = $1/(-\Sigma p_i \log p_i)$

where, $p_i$ is likelihood of the defect to be classified in the main class Mi.

The denominator of the above equation is entropy having a property that when $p_i$ is more uniform, the value becomes larger. Therefore, when the $p_i$ is more uniform, in other words, when the $p_i$ is not

5 different depending on the main classes and cannot be classified into a particular class with high reliability, the entropy becomes higher, and the model likelihood becomes small as a result. This model likelihood is calculated for each model.

10 STEP 3: Calculation of combination likelihood (1206)

Defect class likelihood can be formulated as shown below from the rule learning integration likelihood (STEP 1 in Fig. 12) and the model likelihood (STEP 2 in Fig. 12) under each model:

15 $$P_{cj} = \Sigma_i P(C_i|M_i) P(M_i) \qquad \cdots (1)$$

$C_j$: Classification class,

$P_{cj}$: Likelihood of class $C_j$,

Mi: Classification model i,

$P(C_j|M_i)$: Likelihood of class $C_j$ by classification

20 model $M_i$, and

$P(M_i)$: Likelihood of classification model Mi.

The formula (1) can be explained as a formula that the class likelihood is calculated by plural models and a weighted average according to likeliness

25 (= model likelihood) of the classification model per se is calculated.

Thus, an inherent and optimum classification

model can be provided automatically in response to a defect classification request different depending on each user, and it becomes possible to improve the classification performance.

5     The above description was made on the example assuming the model having the rule-based classification apparatus and the learning type classification apparatus connected hierarchically. But, as shown in Fig. 11, likelihoods 1104 to 1106 calculated by a rule-
10 based classification apparatus 1101 and a learning type classification apparatus 1102 which are parallel to each other or a rule-based classification apparatus 1103 may be combined. In Fig. 11, the rule-based classification apparatuses 1101, 1103 do not classify
15 into a particular class but calculate the likelihoods 1104, 1106 according to the so-called fuzzy logic (fuzzy if-then rule). In other words, the likelihood function configured by the designer is used to calculate the likelihood of each class.
20     By adopting the structure shown in Fig. 11, the instable performance of the learning type classification apparatus when the amount of teach data is little can be resolved. For example, when the amount of teach data is little, the model likelihood of
25 the learning type classification apparatus 1102 is lowered, and the likelihoods 1104, 1106 calculated by the rule-based classification apparatuses 1101, 1103 are regarded as relatively important. Conversely, when

the amount of teach data is large, the model likelihood of the learning type classification apparatus 1102 is increased, and the likelihood 1105 calculated by the learning type classification apparatus 1102 may be

5 regarded as relatively important.

Instable performance of the learning type classification apparatus 1102 when the amount of teach data is variable depending on the classes can be resolved. For example, in a class having a little

10 amount of teach data or a region of the characteristic amount space, the model likelihood of the learning type classification apparatus 1102 is lowered, and the likelihoods 1104, 1106 calculated by the rule-based classification apparatuses 1101, 1103 are regarded as

15 relatively important; conversely, in a class having a large amount of teach data or a region of the characteristic amount space, the model likelihood of the learning type classification apparatus 1102 is raised, and the likelihood 1105 calculated by the

20 learning type classification apparatus 1102 may be regarded as relatively important.

Thus, an inherent and optimum classification model can be provided automatically in response to the defect classification request different depending on

25 the individual users, and it becomes possible to improve the classification performance.

The embodiments of the invention were described above.

In the above description, the classification of defect images detected by the electron type image detector was described. The invention can similarly be applied to the classification of defect images detected

5 by an optical image detector.

Besides, defects may be classified by analyzing output information (e.g., an X-ray spectrum by EDX) from at least one of plural electron or optical type image detectors or a defect analyzing apparatus

10 and calculating a characteristic amount vector of each defect. Of course, the defects may be classified by calculating the characteristic amount vector from the output information alone of the defect analyzing apparatus or the characteristic amount vector may be

15 calculated according to reference data corresponding to the defects.

The invention was described in detail with reference to the examples of the classification of the images of defects caused on the surface of a

20 semiconductor electronic circuit board, but the application of the invention is not limited to the above description.

For example, the invention can also be applied to a method for automatic classification of an

25 image of a defect portion formed on the surface of a printed circuit board, an FPD, a liquid crystal display board, or the like.

According to the present invention, the

classification performance can be improved by automatically providing an inherent and appropriate classification model in response to a defect classification request which is different depending on

5   individual users.

And, according to the present invention, an optimum classification model can be provided automatically to stabilize the classification performance even when the classification of defects requested by the

10   user is changed because of occurrence of a new defect class due to a change in process, or the like.

Besides, according to the present invention, necessity of manual setting of the classification model in response to a defect classification request can be

15   eliminated by automatically providing an optimum classification model.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present

20   embodiment is therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description and all changes which come within the meaning and range of

25   equivalency of the claims are therefore intended to be embraced therein.